Perspectives: Standards and Implementation

ISO 8000 The International Standard for Data Quality

Please describe data profiling.

Data profiling consists of creating formatting, validating and mask rules and some systems support conditional rules (if this field exists, this field should also exist). Profiling does not address data portability or the use of the data.

One of the data quality challenges is to ensure purpose of data use is compatible with purpose of data capture.

Agree 100%, that is why we not only need the technical requirements to set expectations, we need to have the policy and governance frameworks to maintain trust for the exchange of data between sending and receiving systems.

Please specify the difference between data quality and data profiling

ISO 8000 data quality covers data portability and fitness for purpose. Data profiling is focused on detecting potential errors in the data itself.

Great Presentations. Do you have any recommendations on addressing the issue with terms and their coding that get transformed as they travel throughout the health IT ecosystem? Like the telephone game, the meaning of "downstream" can be quite different with data loss along the way. When terms change, the coding changes (if present) to match the provider preferred terms too.

A dictionary helps solve this problem. Data without a dictionary is ambiguous and therefore not portable. If each party has a dictionary (and shares it by publishing it in an open technical dictionary), concept mapping is very straight forward and so very important.

Each Health IT has their own dictionary, and they rename terms based upon the terms in their dictionary (and concept maps too to match with both losing details at times).

A common problem when dictionaries are mismatched on the level of detail of the definition.

Data Usability Implementation Guide

Can you highlight some of what you plan to add in v2 of the sequoia IG—- anything other than FHIR support?

To follow up, some of the v2 possible work items may include event notification, and public health requirements that leverage v2 that align with the modernization efforts. Also under consideration will be v2 focused for the enhanced laboratory guidance provided as well.

Note that support for FHIR is the focus of the development work that was recently kicked off to develop a v2 of the TEFCA Common Agreement and QHIN Technical Framework. This is a DIFFERENT "v2" than the one we are working on within the Sequoia Data Usability initiative.

There are many options, but we will need to determine what really is implementable in another 18 month timeline when the clock resets with the version 2.0 future work underway. The Data Usability Workgroup will be meeting monthly and the Lab Tiger Team will also meet monthly. So please join the work if you feel called.

USCDI Data Quality Assessment

Are we too late in the game? As was stated by one of the panelists, there is no accepted quality standard at this time. AI is now being "trained" on this non-standardized data to make clinical decisions. Is there a way to correct this?

It will be critical for us to retrain our models over time as both the content and the quality of health data advances and as new uses are identified.

USCDI is a fairly loose data set in terms of definitional specificity–- many of its data elements can include a fairly wide variety of data as long as it's the general category covered. Can you speak to how using it as a barometer helps us understand whether the data is specific enough and consistent enough to meet viable data quality standards?

It is true that the specificity of USCDI is variable across the data classes and elements. As we add new data elements to new versions, ONC has set the bar higher than they did/could when they published the first version, built on the CCDS, which had little if any specificity. The USCDI has been designed and evolved to meet many purposes in US health care. It is not perfect for this data application, but it is the best that we have, and as we measure the completeness and mapping of clinical data sets against this standard we can evolve the standard more thoughtfully, including adding specificity to elements that are already included.

Are there any concerns with regards to normalizing or "fixing" of data (with AI or human help) by network intermediaries or QHINs? Identifying and resolving discrepancies in data to ensure meaningful usage is critical, but how do we address concerns with regards to perseverance of data as entered by the provider?

It will always be necessary to maintain the source data and make that available to the data user. There is also value in quality enhanced/remediated data with transparency regarding what data modifications were made, when, by whom/what process, and why. These data transformations should all be considered part of the provenance of the data.

USCDI lacks recognition that data content must be bound to its context. Otherwise it's just a collection of data without known inter-relationships.

Health data is highly imperfect, and has nonetheless been used to inform care, since the first medical records were kept on papyrus. USCDI is but one tool we have to help move us toward the improved access, exchange, and use of health data. The window is open NOW to provide public comment to contribute to the development of USCDI v5 next year. A great opportunity to provide detailed input about how we might capture, reflect, and exchange data regarding Context, as you define this term.

Agree with previous speaker USCDI claims to be useful for any/every purpose. It embodies the problem that if your tool is a hammer, it becomes the proclaimed solution in every instance. Thus, the downstream challenge of rendering useful/usable information at the point of care/point of use.

Thanks for your comment. USCDI comprises a core set of data that must be exchanged by certified health IT modules in support patient care and facilitate patient access using health IT. It establishes a consistent baseline of data for other use cases. USCDI is not an implementation guide or a specification and is implementation standards agnostic.

Perspective: Real World Observations and Best Practices

Measuring Data Quality in Healthcare

Did the patients consent to having their data used for this analysis?

All analyzed data was completely de-identified.

And all data sources agreed to the use of de-identified data in data use agreements.

How was the analysis carried out, can you share the method, tools, software used?

The data was extracted from a static file, serialized into a simple data class canonical model, and processed through a proprietary data quality assessment utility (for data at rest) using USCDI as a criteria with the HDQT applied based on the nature of the assessment. Current versions of the standards were used for conformance relative to RxNorm, LOINC, UCUM and SNOMED CT. The software was an enhanced version of our proprietary PIVOT platform.

Thank you for sharing your approach and findings! Could you say a bit about the differences between the concepts of the "computable" and "complete" measures? In what situation would data be complete but not computable, or vice versa?

Computable (as defined in the HDQT) means that the data is available as discreet information. This typically would apply to a simple attribute. "Complete" means that it has all the parts necessary. The confusion comes from when an attribute checks the 'complete' box by existing - it is 'computable', therefore it is 'complete'. A better example would be a medication concept. It can have computable components, but it is made up of code system, code, and term. If any of those are missing, the medication concept is incomplete. All this being said, it is not perfect and we are looking forward to evolving and refining it.

Is it possible to share measure specifications for some of these (such as for Valid)? Additionally, is data remediation informed by the use of data (that is, were providers or clinical care teams consulted in the creation of these metrics)?

The measure specification is a product of the intersection of the evaluation criteria and the taxonomy. Therefore, the specification for 'Valid' depends on the attribute and how you are evaluating it. For example, a valid numeric result value was a number, any integer or decimal. The issue with many of the labels we use in data quality is the ambiguity and interchangeability of terms. So the way I applied the taxonomy was "valid" meant the attribute met the minimum 'quacks like a duck' evaluation based on the nature of the attribute. With respect to data remediation, for this sample we used high certainty, automated remediation without clinical review. Would not suggest this in a real world scenario. There is automation that can provide high certainty mapping but oversight and review by a learned intermediary is always a good idea.

Thank you. What data field(s) do we use to compute the "Fill Status" calculation - should we assume that this data is coming from the Pharmacy or Claims so by default we know that the prescription was filled? Forgive me if I might have missed that part of the commentary....Great work!

It is my understanding that this data was from C-CDAs and the data used for "fill status" came from a 'status' field that had information that was consistent with the recommendations for fill status in HL7. Different formats have different, but similar, field locations for fill status. The one that is more tricky is the new 'Medication Adherence' attribute specified in USCDI v4. Thanks!

AHIMA Data Quality

ISO 21089 (Health Informatics - Trusted End-to-End Information Flows) is focused on the lifespan and lifecycle of health records at the entry level. If implemented in FHIR, there's a corresponding FHIR R5 Record Lifecycle Event Implementation Guide.

Data quality and including the right content will still be incredibly important.

Patients' maiden names present regular challenges with patient matching/record linkage.

Yes they do, but if you capture as an alternative name. It is minimized.

Even less common names often are fully repeated even within a single health system. So even with a naming convention that provides consistency there are likely to be a lot of name duplicates - and some have a lot of other data duplication across the board. There was an example from a Colorado hospital a few years ago that still had 13 different people with the same first, middle, and last names (that were not what most folks would consider very common) plus three other points of data agreement. Support standardizing name capture, but it seems like the speaker feels like this will solve patient matching and that's very optimistic

To clarify, the AHIMA Naming Framework is a floor that does support improvement in patient matching. Recognizing that as we learn more and explore options - the floor will rise. Incremental improvements bring value. Glad to talk to you about it personally as well.

Are you working with FAST Identity and Matching?

The Framework is for capture of information - accurate content. The SMEs involved in development work are across healthcare. Glad to discuss more if you would like.

If there is no specification for what stands for "fit for use," then anything qualifies under a null specification. This Data Quality information model can be captured, managed in FHIR (unlike CDA) via an IG designed for that purpose. The HL7 Reducing Clinician Burden WG under HL7 EHR Systems WG welcomes queries on this work.

Inclusion of the right content improves data quality. Inclusion of the end-user in the process will help identify "fit for use."

Some causes for poor clinical data quality are:

- Flaws in Clinical Data Frameworks:
- Imperfect Clinical Processes: Clinical processes vary widely across healthcare institutions. These variations can lead to inconsistencies in data collection and interpretation.
- Over-reliance on Manual Data Entry: Many clinical data systems rely heavily on manual data entry by clinicians. This can lead to errors, omissions, and inconsistencies in the data.
- Lack of Standardization: Without standardized data entry guidelines, the quality and consistency of data can vary significantly. This can make data analysis and interpretation challenging.
- Workflow Disruptions: Implementing a new data framework can disrupt existing clinical workflows, leading to resistance from clinicians and staff. Improves transparency on how a predictive DSI was designed, developed, trained, evaluated, and should be used, addressing fundamental information asymmetries in the marketplace for predictive DSIs.

Veteran's Health Information Exchange (VHIE)

One challenge is we are running into is the immense amount of work that needed for mapping discrete data to the correct standard. Any recommendations on how to work through this.

It is clear that we cannot effectively and scalably do this work exclusively with manual human effort. A number of groups are working to leverage AI/ML to assist with this. Initial results seem to be positive, but these are early days.

With TEFCA at the forefront of interoperabilty, will there be a standard of data usability (discrete data) in order to use this data to meet things like quality measures?

Adding data quality and/or usability standards into the TEFCA framework is being discussed. I personally feel that we should be advocating for this addition.

We have focused too long on data standards for exchange at the back-end. These standards essentially demand transformation of data content/context from its source representation to the exchange artifact (e.g., HL7 v2 message, CDA document, FHIR resource) to the receiver representation. Two transformations in the course of a single exchange instance. Transformation often introduces alterations, omissions and errors in data content/context. And with it goes data quality, fidelity to source access, as FHIR is now mandated through the 21st Century Cures Act, and as the first standard data model we've had in the US, my recommendation to my clients is to map legacy data to FHIR. What's important to remember that this is not just a technical exercise, but mapping needs to ensure it is apples to apples conceptually. My two cents. Would love to hear others' thoughts on this as well.

Those use cases seem very wide - how do you determine the level of granularity that's reasonable for defining purpose or some of these other concepts so that they're actually useful?

The TEFCA RCE is developing detailed "standard operating procedures" to specify the sub-use cases that will be supported by TEFCA initially, and these will be advanced over time. See: https://rce.sequoiaproject.org/tefca-and-rce-resources/

We need more clinicians with formal training in informatics.

Agreed; formal training in informatics is needed across the spectrum of clinical disciplines and specialties. Each community has unique needs.

It would also be helpful if, as an industry, there was more understanding of what Clinical Informatics is and what they are trained to do. Many people do not know what a CI is or the skill sets they have.

This is a health care issue. As someone who worked in tech in many other industries, the lack of real standards drives me absolutely insane. It would not be tolerated in any other industry. Ditto lack of backwards compatibility.